PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE
Nouvelle série, tome 83(97) (2008), 27–35 DOI: 10.2298/PIM0897027O

SINGLES IN A MARKOV CHAIN

Edward Omey and Stefan Van Gulck

Communicated by Slobodabka Janković

Abstract. Let $\{X_i, i \geqslant 1\}$ denote a sequence of variables that take values in $\{0,1\}$ and suppose that the sequence forms a Markov chain with transition matrix P and with initial distribution $(q,p) = (P(X_1 = 0), P(X_1 = 1))$. Several authors have studied the quantities S_n, $Y(r)$ and $AR(n)$, where $S_n = \sum_{i=1}^{n} X_i$ denotes the number of successes, where $Y(r)$ denotes the number of experiments up to the r-th success and where $AR(n)$ denotes the number of runs. In the present paper we study the number of singles $AS(n)$ in the vector $(X_1, X_2, \ldots, X_n)$. A single in a sequence is an isolated value of 0 or 1, i.e., a run of length 1. Among others we prove a central limit theorem for $AS(n)$.

1. Introduction

Many papers are devoted to sequences of Bernoulli trials and they form the basis of many (known) distributions and scientific activities. Applications are numerous. To mention only a few:

– the one-sample runs test can be used to test the hypothesis that the order in a sample is random;

– the number of successes can be used for testing for trends in the weather or in the stock market;

– Bernoulli-trials are important in matching DNA-sequences;

– the number of (consecutive) failures can be used in quality control.

In the case where the trials are i.i.d. many results are known concerning e.g. the quantities S_n, $Y(r)$ and $AR(n)$, where $S_n = \sum_{i=1}^{n} X_i$ denotes the number of successes, where $Y(r)$ denotes the number of experiments up to the r-th success and where $AR(n)$ denotes the number of runs. A Markovian binomial distribution and other generalizations of the binomial distribution was studied e.g. by Altham [**1**], Madsen [**7**], Omey et al. [**8**]. In the present paper we study the number of singles $AS(n)$ in the vector $(X_1, X_2, \ldots, X_n)$.

Suppose that each X_i takes values in the set $\{0,1\}$ and for $n \geqslant 1$, let $AS(n)$ denote the number of singles in the sequence $(X_1, X_2, \ldots, X_n)$. With $AS(n)$ we

2000 *Mathematics Subject Classification*: Primary 60J10; Secondary 60F05, 60K99, 60J20.

count the number of isolated values of 0 or 1 in $(X_1, X_2, \dots, X_n)$. Mathematically we can study $AS(n)$ as follows. For fixed $n \geqslant 1$ we construct a new sequence of $t\{0,1\}$-valued r.v. t_i where $t_i = 1$ if and only if X_i is a single. More precisely we define the t_i as follows:

$$t_1 = (X_2 - X_1)^2, \quad t_n = (X_n - X_{n-1})^2;$$
$$t_i = (X_{i+1} - X_i)^2 (X_i - X_{i-1})^2, \quad 2 \leqslant i \leqslant n-1.$$

Clearly we have $AS(n) = \sum_{i=1}^{n} t_i$. Note that for simplicity we use the notation t_i and not the notation $t_i^{(n)}$. In studying t_i and $AS(n)$ we assume that the sequence $X_1, X_2, \dots, X_n, \dots$ is a Markov chain taking values in $\{0,1\}$. As special cases we recover the i.i.d. case. We also briefly consider the the number of 0-singles $AS_n^{(0)}$ and the number of 1-singles $AS_n^{(1)}$, i.e., $AS_n^{(0)}$ counts the number of isolated zeros in the sequence and $AS_n^{(1)}$ counts the number of isolated "1" in the sequence.

Before starting our analysis we briefly discuss the Markov chain we use. We assume that $\{X_i, i \geqslant 1\}$ is a $\{0,1\}$-Markov chain with initial distribution

$$(P(X_1 = 0), P(X_1 = 1)) = (q, p), \quad \text{where } 0 < p = 1 - q < 1.$$

The transition matrix P is given by

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{pmatrix}$$

where for $i, j = 0, 1$, $p_{i,j} = P(X_2 = j \mid X_1 = i)$. To avoid trivialities we suppose that $0 < p_{i,j} < 1$. Note that the Markov chain has the unique stationary vector given by $(x, y) = (p_{1,0}, p_{0,1})/(p_{0,1} + p_{1,0})$. The eigenvalues of P are given by $\lambda_1 = 1$ and $\lambda = 1 - p_{0,1} - p_{1,0} = p_{0,0} - p_{1,0}$. Note that $|\lambda| < 1$. By induction it is easy to show that the n-step transition matrix is given by

$$P^n = A + \lambda^n B, \text{ where } A = \begin{pmatrix} x & y \\ x & y \end{pmatrix} \text{ and } B = \begin{pmatrix} y & -y \\ -x & x \end{pmatrix}. \tag{1.1}$$

Using these relations we find that

$$\big(P(X_n = 0), P(X_n = 1)\big) = (q, p)P^{n-1} = \big(x + \lambda^{n-1}(y - p), y - \lambda^{n-1}(y - p)\big).$$

Among others this implies (see Omey et al. [**8**]) that for $n \geqslant 1$ we have

$$E(X_n) = y - \lambda^{n-1}(y - p),$$
$$\operatorname{Var}(X_n) = \big(y - \lambda^{n-1}(y - p)\big)\big(x + \lambda^{n-1}(y - p)\big),$$
$$\operatorname{Cov}(X_m, X_n) = \lambda^{n-m} \operatorname{Var}(X_m), \quad m \leqslant n.$$

As a special case we consider the case where the transition matrix $P = P(p, \rho)$ is given by

$$P(p, \rho) = \begin{pmatrix} q + \rho p & p(1 - \rho) \\ q(1 - \rho) & p + \rho q \end{pmatrix}.$$

In this case we have $(x, y) = (q, p)$ and $\lambda = \rho$. Since we also have $P(X_n = 1) = p$, for all n, the X_i have the same distribution. If $\rho \neq 0$, the X_i are correlated with $\rho = \rho(X_n, X_{n+1})$. From this it follows that $\operatorname{Cov}(X_n, X_m) = \rho^{n-m} pq$ $(m \leqslant n)$. This type of correlated Bernoulli trials has been studied among others by Dimitrov and Kolev [**3**]. See also Kupper and Haseman [**5**] or Lai et al. [**6**]. If $\rho = 0$, we find back the case where the X_i are i.i.d. Bernoulli variables. In Fu and Lou [**4**], the authors use a finite Markov imbedding approach to study runs and patterns.

2. Moments

Now we focus our attention on the number of singles. We use the sequence of r.v. t_i as in the introduction. In Propositions 2.1 and 2.2 below we study distributional properties of the random variables t_i.

PROPOSITION 2.1. *For $n \geqslant 3$ we have:*

- $P(t_1 = 1) = pp_{1,0} + qp_{0,1}$;
- $P(t_i = 1) = p_{0,1}p_{1,0}$, *for* $2 \leqslant i \leqslant n-1$;
- $P(t_n = 1) = P(X_{n-1} = 1)p_{1,0} + P(X_{n-1} = 0)p_{0,1} = (q,p)P^{n-2}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}$.

PROOF. For t_1 we have

$$P(t_1 = 1) = P\big((X_1, X_2) \in \{(1,0),(0,1)\}\big) = pp_{1,0} + qp_{0,1}.$$

For $2 \leqslant i \leqslant n-1$, we have

$$P(t_i = 1) = P\big((X_{i-1}, X_i, X_{i+1}) \in \{(0,1,0),(1,0,1)\}\big)$$

and it follows that

$$P(t_i = 1) = \big(P(X_{i-1} = 0) + P(X_{i-1} = 1)\big)p_{0,1}p_{1,0} = p_{0,1}p_{1,0}.$$

Finally, we have $P(t_n = 1) = P\big((X_{n-1}, X_n) \in \{(1,0),(0,1)\}\big)$ so that

$$P(t_n = 1) = P(X_{n-1} = 1)p_{1,0} + P(X_{n-1} = 0)p_{0,1}. \qquad \square$$

PROPOSITION 2.2. *For $n \geqslant 4$, the joint distributions are given by:*

(a) *For $i = 1$ or $i = n-1$, $P(t_i = t_{i+1} = 1) = p_{0,1}p_{1,0}$.*

(b) *For $2 \leqslant i \leqslant n-2$, $P(t_i = t_{i+1} = 1) = p_{0,1}p_{1,0}(q,p)P^{i-2}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}$.*

(c) $P(t_1 = t_n = 1) = (pp_{1,0}, qp_{0,1})P^{n-3}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}$.

(d) *For $2 \leqslant i \leqslant n-2$, $P(t_i = t_n = 1) = p_{0,1}p_{1,0}(q,p)P^{n-4}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}$.*

(e) *In all other cases t_i and t_j are independent.*

PROOF. (a) For (t_1, t_2) we have

$$P(t_1 = t_2 = 1) = P\big((X_1, X_2, X_3) \in \{(1,0,1),(0,1,0)\}\big)$$

so that $P(t_1 = t_2 = 1) = pp_{1,0}p_{0,1} + qp_{0,1}p_{1,0} = p_{1,0}p_{0,1}$. The result for $i = n-1$ follows in a similar way.

(b) For $i = 2, 3, \ldots, n-2$ we have

$$P(t_i = t_{i+1} = 1) = P\big((X_{i-1}, X_i, X_{i+1}, X_{i+2}) \in \{(1,0,1,0),(0,1,0,1)\}\big)$$

so that

$$P(t_i = t_{i+1} = 1) = P(X_{i-1} = 1)p_{1,0}p_{0,1}p_{1,0} + P(X_{i-1} = 0)p_{0,1}p_{1,0}p_{0,1}.$$

Using $\big(P(X_{i-1}=0), P(X_{i-1}=1)\big) = (q,p)P^{i-2}$ we find that

$$P(t_i = 1, t_{i+1} = 1) = p_{0,1}p_{1,0}(q,p)P^{i-2}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}.$$

(c) For (t_1, t_n) we have $P(t_1 = t_n = 1) = P\big((X_1, X_2, X_{n-1}, X_n) \in S\big)$ where $S = \{(1,0,1,0), (1,0,0,1), (0,1,1,0), (0,1,0,1)\}$. Considering the first case, we have

$$P\big((X_1, X_2, X_{n-1}, X_n) = (1,0,1,0)\big) = pp_{1,0}p_{0,1}^{(n-3)}p_{1,0}.$$

In a similar way we calculate the other 3 cases. Using matrices, it follows that

$$P(t_1 = t_n = 1) = (pp_{1,0}, qp_{0,1})P^{n-3}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}.$$

(d) For $2 \leqslant i \leqslant n-2$ we have

$$P(t_i = t_n = 1) = P\big((X_{i-1}, X_i, X_{i+1}, X_{n-1}, X_n) \in S\big)$$

where $S = \{(1,0,1,0,1), (1,0,1,1,0), (0,1,0,0,1), (0,1,0,1,0)\}$. Considering the first case, we have

$$P\big((X_{i-1}, X_i, X_{i+1}, X_{n-1}, X_n) = (1,0,1,0,1)\big) = P(X_{i-1} = 1)p_{1,0}p_{0,1}p_{1,0}^{(n-i-2)}p_{0,1}.$$

In a similar way we treat the other cases and using matrices we find that

$$P(t_i = t_n = 1) = p_{0,1}p_{1,0}\big(P(X_{i-1}=0), P(X_{i-1}=1)\big)P^{n-i-2}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}$$

so that

$$P(t_i = t_n = 1) = p_{0,1}p_{1,0}(q,p)P^{n-4}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}.$$

(e) To prove independence, consider for example (t_1, t_3). We have

$$P(t_1 = t_3 = 1) = P\big((X_1, X_2, X_3, X_4) \in \{(1,0,1,0), (0,1,0,1)\}\big)$$

so that

$$P(t_1 = t_3 = 1) = pp_{1,0}p_{0,1}p_{1,0} + qp_{0,1}p_{1,0}p_{0,1} = P(t_1 = 1)\,P(t_3 = 1).$$

It follows that t_1 and t_3 are independent. In a similar way it follows that (t_1, t_i) for $i = 3, 4, \ldots, n-1$ are independent r.v. and that the other (t_i, t_j) are independent r.v. □

In the i.i.d. case, we obtain the following corollary.

Corollary 2.1. *Suppose $n \geqslant 4$ and $X_1, X_2, \ldots, X_n$ i.i.d. with $P(X_1 = 1) = p$; then*

(a) $P(t_1 = 1) = P(t_n = 1) = 2pq$ *and for* $2 \leqslant i \leqslant n-1$, $P(t_i = 1) = pq$.
(b) $P(t_1 = t_2 = 1) = P(t_{n-1} = t_n = 1) = pq$ *and for* $2 \leqslant i \leqslant n-2$, $P(t_i = t_{i+1}) = 2p^2q^2$.
(c) $P(t_1 = t_n = 1) = 4p^2q^2$.
(d) *For* $2 \leqslant i \leqslant n-2$, $P(t_i = t_n = 1) = 2p^2q^2$.
(e) *In the other cases t_i and t_j are independent.*

In the next result we discuss the mean and the variance of $AS(n)$.

Proposition 2.3. (a) *As* $n \to \infty$, *we have* $\frac{1}{n} E(AS(n)) \to p_{0,1}p_{1,0}$.

(b) *As* $n \to \infty$, *we have* $\frac{1}{n}\operatorname{Var}(AS(n)) \to p_{0,1}p_{1,0}\left(1 - 3p_{0,1}p_{1,0} + \dfrac{4p_{0,1}p_{1,0}}{p_{0,1}+p_{1,0}}\right)$.

Proof. (a) Using Proposition 2.1, for $2 \leqslant i \leqslant n-1$ we have $E(t_i) = p_{0,1}p_{1,0}$ It follows that $E(AS(n)) = (n-2)p_{0,1}p_{1,0} + E(t_1) + E(t_n)$ and the result follows.

(b) Using Proposition 2.2 we have

$$\operatorname{Var}(AS(n)) = \sum_{1=1}^{n} \operatorname{Var}(t_i) + 2\sum_{i=1}^{n-2} \operatorname{Cov}(t_i, t_{i+1}) + 2\sum_{i=1}^{n-1} \operatorname{Cov}(t_i, t_n) = I + II + III.$$

We consider these three terms separately.

Term I. For $i = 2, 3, , \ldots, n-1$ we have $\operatorname{Var}(t_i) = p_{0,1}p_{1,0}(1 - p_{0,1}p_{1,0})$. For $i = 1, n$, we have $\operatorname{Var}(t_1) + \operatorname{Var}(t_n) \leqslant 2$. It follows that $I/n \to p_{0,1}p_{1,0}(1 - p_{0,1}p_{1,0})$.

Term II. For $i = 2, 3, \ldots, n-2$ it follows from Propositions 2.1 and 2.2 that

$$\operatorname{Cov}(t_i, t_{i+1}) = p_{0,1}p_{1,0}(q,p)P^{i-2}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix} - (p_{0,1}p_{1,0})^2.$$

It follows that

$$\sum_{i=2}^{n-2} \operatorname{Cov}(t_i, t_{i+1}) = p_{0,1}p_{1,0}(q,p)\sum_{i=2}^{n-2} P^{i-2}\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix} - (n-3)(p_{0,1}p_{1,0})^2.$$

Using $P^k = A + \lambda^k B$, cf (1.1), we obtain that

$$\frac{1}{n}\sum_{i=2}^{n-2} P^{i-2} = \frac{1}{n}\sum_{j=0}^{n-4}(A + \lambda^j B) \to A.$$

We conclude that

$$\frac{II}{n} \to 2p_{0,1}p_{1,0}(q,p)A\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix} - 2(p_{0,1}p_{1,0})^2 = 2p_{0,1}p_{1,0}(xp_{0,1} + yp_{1,0} - p_{1,0}p_{0,1}).$$

Term III. For $2 \leqslant i \leqslant n-1$, we have

$$\operatorname{Cov}(t_i, t_n) = p_{0,1}p_{1,0}(q,p)\left(P^{n-4} - P^{n-2}\right)\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}$$

so that

$$\operatorname{Cov}(t_i, t_n) = p_{0,1}p_{1,0}(q,p)\left(\lambda^{n-4} - \lambda^{n-2}\right)B\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix},$$

$$\sum_{i=2}^{n-1} \operatorname{Cov}(t_i, t_n) = (n-3)p_{0,1}p_{1,0}(q,p)\left(\lambda^{n-4} - \lambda^{n-2}\right)B\begin{pmatrix} p_{0,1} \\ p_{1,0} \end{pmatrix}.$$

It follows that $III/n \to 0$. We conclude that

$$\frac{1}{n}\operatorname{Var}(AS(n)) \to p_{0,1}p_{1,0}(1 - p_{1,0}p_{0,1}) + 2p_{0,1}p_{1,0}(xp_{0,1} + yp_{1,0} - p_{1,0}p_{0,1}).$$

Using $x = p_{1,0}/(p_{1,0}+p_{0,1})$ and $y = 1-x$, it follows that

$$\frac{1}{n}\operatorname{Var}(AS(n)) \to p_{0,1}p_{1,0}\left(1-3p_{0,1}p_{1,0}+\frac{4p_{0,1}p_{1,0}}{p_{0,1}+p_{1,0}}\right). \qquad \square$$

REMARK 2.1. A more detailed analysis shows that

$$E(AS(n)) = (p+y)p_{1,0}+(q+x)p_{0,1}+(n-2)p_{0,1}p_{1,0}+\lambda^{n-2}(y-p)(p_{0,1}-p_{1,0}).$$

In the next corollary we formulate two special cases.

COROLLARY 2.2. (i) *If* $P = P(p,\rho)$ *then*

$$E(AS(n)) = (n-2)pq(1-\rho)^2+4pq(1-\rho),$$

$$\frac{1}{n}\operatorname{Var}(AS(n)) \to pq(1-\rho)^2\big(1+pq(1-\rho)+3pq\rho(1-\rho)\big).$$

(ii) (*the i.i.d. case*) *If* $\rho = 0$, *then*

$$E(AS(n)) = (n+2)pq, \quad \text{and} \quad \frac{1}{n}\operatorname{Var}(AS(n)) \to pq(1+pq).$$

3. The distribution of $AS(n)$

In the next proposition we show how to calculate $p_n(k) = P(AS(n) = k)$ recursively.

For $n \geqslant 2$ and for $i,j = 0,1$ we write

$$p_n(k) = \sum_{i=0}^{1}\sum_{j=0}^{1} p_n^{(i,j)}(k), \quad \text{where} \quad p_n^{(i,j)}(k) = P(AS(n) = k, X_{n-1} = i, X_n = j)$$

For $n = 2$ we clearly have $p_2^{(0,0)}(0) = qp_{0,0}$ and 0 otherwise; also $p_2^{(0,1)}(2) = qp_{0,1}$ and 0 otherwise; $p_2^{(1,0)}(2) = pp_{1,0}$ and 0 otherwise and $p_2^{(1,1)}(0) = pp_{1,1}$ and 0 otherwise. We have the following relations.

PROPOSITION 3.1. *For* $n \geqslant 2$ *we have*

- $p_{n+1}^{(0,0)}(k) = p_{0,0}p_n^{(1,0)}(k+1) + p_{0,0}p_n^{(0,0)}(k)$;
- $p_{n+1}^{(0,1)}(k) = p_{0,1}p_n^{(0,0)}(k-1) + p_{0,1}p_n^{(1,0)}(k-1)$;
- $p_{n+1}^{(1,0)}(k) = p_{1,0}p_n^{(0,1)}(k-1) + p_{1,0}p_n^{(1,1)}(k-1)$;
- $p_{n+1}^{(1,1)}(k) = p_{1,1}p_n^{(0,1)}(k+1) + p_{1,1}p_n^{(1,1)}(k)$.

PROOF. We only prove the first relation. We have $p_{n+1}^{(0,0)}(k) = I + II$ where

$$I = P(AS(n+1) = k,\ X(n-1) = 0,\ X(n) = 0,\ X(n+1) = 0)$$
$$II = P(AS(n+1) = k,\ X(n-1) = 1,\ X(n) = 0,\ X(n+1) = 0).$$

It follows that

$$\begin{aligned} I &= P\big(AS(n) = k, X(n-1) = 0, X(n) = 0, X(n+1) = 0\big) \\ &= P\big(X(n+1) = 0 \mid X(n) = 0, X(n-1) = 0, AS(n) = k\big)p_n^{0,0}(k) \end{aligned}$$

so that $I = p_{0,0}p_n^{(0,0)}(k)$.

In a similar way we have $II = p_{0,0}p_n^{(1,0)}(k+1)$. □

Proposition 3.1 can be used to calculate the p.d. of $AS(n)$ explicitly for small values of n. A straightforward analysis shows that the complexity effort is of order n^2 and exact calculations can be carried out for moderate values of n. For large values of n we prove the following central limit theorem.

THEOREM 3.1. *As $n \to \infty$, we have*

$$\frac{AS(n) - np_{0,1}p_{1,0}}{\sqrt{n}} \overset{d}{\Rightarrow} Z,$$

where $Z \sim N(0,\beta)$ with $\beta = p_{0,1}p_{1,0}\left(1 - 3p_{0,1}p_{1,0} + 4\dfrac{p_{0,1}p_{1,0}}{p_{1,0}+p_{0,1}}\right)$.

PROOF. To prove the result we use Proposition 3.1 and generating functions. Let $\Psi_n^{(i,j)}(z)$ denote the generating function of $p_n^{(i,j)}(k)$ and let $\Psi_n(z)$ denote the generating function of $p_n(k)$. Also, let

$$\Lambda_n(z) = \left(\Psi_n^{(0,0)}(z), \Psi_n^{(0,1)}(z), \Psi_n^{(1,0)}(z), \Psi_n^{(1,1)}(z)\right).$$

Clearly we have

$$\Lambda_2(z) = (qp_{0,0}, qp_{0,1}z^2, pp_{1,0}z^2, pp_{1,1}) \quad \text{and} \quad \Psi_n(z) = \Lambda_n(z)(1,1,1,1)^t.$$

For $n \geqslant 2$ we use Proposition 3.1 to see that

- $\Psi_{n+1}^{(0,0)}(z) = (p_{0,0}/z)\Psi_n^{(1,0)}(z) + p_{0,0}\Psi_n^{(0,0)}(z)$;
- $\Psi_{n+1}^{(0,1)}(z) = p_{0,1}z\Psi_n^{(0,0)}(z) + p_{0,1}z\Psi_n^{(1,0)}(z)$;
- $\Psi_{n+1}^{(1,0)}(z) = p_{1,0}z\Psi_n^{(0,1)}(z) + p_{1,0}z\Psi_n^{(1,1)}(z)$;
- $\Psi_{n+1}^{(1,1)}(z) = (p_{1,1}/z)\Psi_n^{(0,1)}(z) + p_{1,1}\Psi_n^{(1,1)}(z)$.

For $n \geqslant 2$ we obtain that $\Lambda_{n+1}(z) = \Lambda_n(z)A(z) = \Lambda_2(z)A^{n-1}(z)$, where the matrix $A(z)$ is given by

$$A(z) = \begin{pmatrix} p_{0,0} & p_{0,1}z & 0 & 0 \\ 0 & 0 & p_{1,0}z & p_{1,1}/z \\ p_{0,0}/z & p_{0,1}z & 0 & 0 \\ 0 & 0 & p_{1,0}z & p_{1,1} \end{pmatrix}.$$

The eigenvalue equation of $A(z)$ leads to

$$\lambda^4 - \lambda^3(p_{0,0}+p_{1,1}) + \lambda^2(p_{0,0}p_{1,1} - z^2p_{0,1}p_{1,0}) - \lambda a(z) - b(z) = 0, \tag{3.1}$$

where

$$a(z) = z(1-z)(p_{0,0}+p_{1,1})p_{0,1}p_{1,0}, \quad \text{and} \quad b(z) = p_{0,0}p_{0,1}p_{1,0}p_{1,1}(1-z)^2.$$

In the case where $z = 1$ the eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 1 - p_{0,1} - p_{1,0}$ and $\lambda_3 = \lambda_4 = 0$. In the general case, a continuity argument shows that for $z < 1$, the matrix $A(z)$ has a unique largest eigenvalue $\lambda(z) = \lambda_1(z)$ such that $\lambda(z) \to 1$ as $z \to 1$. The other eigenvalues are dominated by $\lambda(z)$. It follows that

$$A^n(z_n)/(\lambda(z_n))^n \to U(1)$$

where $z_n \to 1$ and where each row of $U(1)$ equals $(xp_{0,0}, xp_{0,1}, yp_{1,0}, yp_{1,1})$.

Now we consider the largest eigenvalue $\lambda(z)$ of $A(z)$. Starting from (3.1) we calculate the first derivative, then the second derivative and then take $z = 1$. Some lengthy but straighforward calculations show that

$$\lambda'(1) = p_{1,0}p_{0,1} \quad \text{and} \quad \lambda''(1) = \frac{2p_{0,1}^2 p_{1,0}^2 (p_{0,0} + p_{1,1})}{p_{0,1} + p_{1,0}}.$$

Using Taylor's expansion for $\log(z)$ and for $\log(\lambda(z))$ around $z = 1$, we find that

$$\frac{\log(\lambda(z)) - p_{1,0}p_{0,1}\log(z)}{(1-z)^2} \to \frac{1}{2}\beta$$

where

$$\beta = p_{0,1}p_{1,0}\left(1 - 3p_{0,1}p_{1,0} + 4\frac{p_{0,1}p_{1,0}}{p_{1,0} + p_{0,1}}\right).$$

Now we replace z by $u_n = z^{1/\sqrt{n}}$ to see that

$$\frac{\lambda^n(u_n)}{u_n^{np_{0,1}p_{1,0}}} \to \exp\Big(\frac{1}{2}\beta(\log(z))^2\Big).$$

Turning to $\Psi_{n+1}(z)$ we find that $\Psi_{n+1}(u_n) \sim \Lambda_2(u_n)\lambda^n(u_n)U(1)(1,1,1,1)^t$ and hence

$$\Psi_{n+1}(u_n)u_n^{-np_{0,1}p_{1,0}} \to \Big(\exp\Big\{\frac{1}{2}\beta(\log(z))^2\Big\}\Big)\Lambda_2(1)\,U(1)\,(1,1,1,1)^t.$$

It follows that

$$\Psi_{n+1}(u_n)u_n^{-np_{0,1}p_{1,0}} \to \exp\Big\{\frac{1}{2}\beta(\log(z))^2\Big\}.$$

Since $\Psi_{n+1}(u_n) = E\big(z^{AS(n+1)/\sqrt{n}}\big)$ the desired result follows. □

In the i.i.d. case we find back the following result of Bloom [**2**].

Corollary 3.1. *In the i.i.d. case we have*

$$\frac{AS(n) - npq}{\sqrt{n}} \overset{d}{\Rightarrow} Z,$$

where $Z \sim N\big(0, \beta = pq(1 + pq)\big)$.

4. Singles "0" and singles "1"

In this section we briefly discuss the number $AS_n^{(0)}$ of isolated values 0 and the number $AS_n^{(1)}$ of isolated values 1. First we look at isolated values of 0. Starting from the sequence $X_1, X_2, \ldots, X_n$ we define $t_i^{(0)} = 1$ if $X_i = 0$ is a single. Clearly we have

$$t_1^{(0)} = X_2(1 - X_1), \quad t_n^{(0)} = X_{n-1}(1 - X_n), \quad t_i^{(0)} = X_{i-1}(1 - X_i)X_{i+1}$$

and $AS_n^{(0)} = \sum_{i=1}^n t_i^{(0)}$. Using the methods of the previous sections one can prove the following result.

THEOREM 4.1. (a) *As* $n \to \infty$ *we have* $\frac{1}{n}E(AS_n^{(0)}) \to yp_{0,1}p_{1,0}$ *and*

$$\frac{1}{n}\operatorname{Var}(AS_n^{(0)}) \to \theta_0 = yp_{0,1}p_{1,0}(1 - 3yp_{0,1}p_{1,0} + 2xyp_{1,0}).$$

(b) *As* $n \to \infty$ *we have*

$$\frac{AS_n^{(0)} - nyp_{0,1}p_{1,0}}{\sqrt{n}} \overset{d}{\Rightarrow} Z^{(0)}$$

where $Z^{(0)} \sim N(0, \theta_0)$.

An entirely similar result holds for $AS_n^{(1)}$. Now we find

$$\frac{AS_n^{(1)} - nxp_{0,1}p_{1,0}}{\sqrt{n}} \overset{d}{\Rightarrow} Z^{(1)}$$

where $Z^{(1)} \sim N(0, \theta_1)$ with $\theta_1 = xp_{0,1}p_{1,0}(1 - 3xp_{0,1}p_{1,0} + 2xyp_{0,1})$. Using

$$\operatorname{Var}(AS(n)) = \operatorname{Var}(AS_n^{(0)}) + \operatorname{Var}(AS_n^{(1)}) + 2\operatorname{Cov}(AS_n^{(0)}, AS_n^{(1)}),$$

we obtain the following asymptotic expression for the covariance.

COROLLARY 4.1. *As* $n \to \infty$ *we have*

$$\frac{1}{n}\operatorname{Cov}\big(AS_n^{(0)}, AS_n^{(1)}\big) \to -3xyp_{0,1}^2p_{1,0}^2 + \frac{2p_{0,1}^2p_{1,0}^2}{p_{1,0} + p_{0,1}}(1 - xy).$$

References

[1] P. Altham, *Two generalizations of the binomial distribution*, J. R. Stat. Soc., Ser. C **27** (1978), 162–167.

[2] D. M. Bloom, *Singles in a sequence of coin tosses*, Coll. Math. J. **29**(2) (1998), 120–127.

[3] B. Dimitrov, B. and N. Kolev, *Extended in time correlated Bernoulli trials in modeling waiting times under periodic environmental conditions*, Technical paper, Universidade de Sao Paulo, Brasil, 1999.

[4] J. C. Fu and W. Y. W. Lou, *Distribution theory of runs and patterns and its applications. A finite Markov chain imbedding approach*, World Scientific, Singapore, 2003.

[5] L. L. Kupper and J. K. Haseman, *The use of the correlated binomial model for the analysis of certain toxicological experiments*, Biometrics **34** (1978), 69–76.

[6] C. D. Lai, K. Govindaraju, and M. Xie, *Effects of correlation on fraction non-conforming statistical process control procedures*, J. Appl. Stat. **25**(4) (1998), 535–543.

[7] R. W. Madsen, *Generalized binomial distributions*, Commun. Stat., Theory Methods **22**(11) (1993), 3065–3086.

[8] E. Omey, J. Santos, and S. Van Gulck, *A Markov binomial distribution*, Appl. Anal. Discrete Math. **2**(1) (2008), 38–50. (available electronically at: `http://pefmath.etf.bg.ac.yu`)

Department of Mathematics and Statistics
HUB, Stormstraat 2, 1000 Brussels, Belgium
and
Affiliated Researcher K. U. Leuven – ORSTAT
Naamsestraat 69, 3000 Leuven, Belgium
edward.omey@hubrussel.be

(Received 06 05 2008)

Department of Mathematics and Statistics
HUB, Stormstraat 2, 1000 Brussels, Belgium
stefan.vangulck@hubrussel.be